Modal Specifications of Trace-Based Security Properties

Rafael Accorsi David Basin Luca Viganò
Institut für Informatik, Albert-Ludwigs-Universität Freiburg
Georges-Köhler-Allee 52, D-79110 Freiburg, Germany
{accorsi,basin,luca}@informatik.uni-freiburg.de

ABSTRACT

We introduce a multi-modal logic that combines complementary features of authentication logics and trace-based approaches. Our logic contains two kinds of modalities: implicit belief, which formalizes the view of an external agent reasoning about interleaved protocol executions, and explicit belief, which uses awareness to model the resource-bounded reasoning of the agents involved in the executions. We employ these modalities to formalize extensional and intensional specifications of protocols and their properties, and use these formalizations to characterize and reason about attacks. As an example, we consider the Needham-Schroeder Public Key protocol and use our logic to demonstrate the existence of the well-known man-in-the-middle attack, and also show the equivalence of our modal specification to one based on an interleaved trace semantics.

1. INTRODUCTION

Security protocols describe how agents should exchange messages to achieve security goals such as confidentiality and integrity of data, or authentication of the identity of agents in a network. A number of approaches have been proposed for rigorously analyzing security protocols. Some of these are based on specialized security logics, such as the foundational BAN logic for authentication protocols [6] and its extensions, e.g. [1, 4, 7, 13, 20, 21]. These logics work by formalizing the doxastic or epistemic reasoning of agents executing a protocol, and security properties are formalized and reasoned about in terms of the way the beliefs or the knowledge of the agents evolve as messages are exchanged. Although effective for finding some kinds of flaws, the logics' semantics are often lacking or restrictive (e.g. the logics are designed to reason about a *single* protocol execution).

An alternative way of reasoning about security protocols is to consider protocols as sets of possibly interleaved communication traces. For example, given a protocol and an attacker model, Paulson [18] turns these into an inductive definition (of the trace set) in higher-order logic. The resulting set can be used to inductively establish security properties by showing that they hold for all traces. Similar inductive definitions are used by Basin in [3] to provide a basis for finding attacks (traces violating security properties) using infinite-state model-checking. The strengths and weaknesses of trace-based methods are in some sense complementary to security logics. Although trace-based methods provide a simple and expressive theory for formalizing the semantics of protocols and security properties in terms of interleaved executions, characterizing attacks as properties of traces can be tricky, whereas BAN-like specifications are generally simpler and more abstract.

In this paper, we introduce a multi-modal logic that combines complementary features of authentication logics and trace-based approaches. Our account is semantic: traces are used to build a Kripke structure upon which a modal logic is defined to reason about actions occurring in interleaved protocol executions. The modalities are used to formalize the *implicit* and *explicit beliefs* of agents, allowing modal specifications of security properties while being based on an underlying interleaved trace semantics. The two belief modalities give us considerable flexibility in our specifications. Roughly speaking, using the implicit belief modality we can model what agents would believe had they seen all messages exchanged by all agents, and using explicit belief we can model what agents believe based on what they have actually seen. We define implicit belief as the standard modality of belief logics [14]. To formalize explicit belief and express the local reasoning of an agent based on the actions he has participated in, we adapt the notion of *awareness* (introduced in logics for artificial intelligence [10, 11, 22] to address the problem of agents having unbounded reasoning power and thus being logically omniscient, which is a characteristic of most doxastic/epistemic logics).

We present two applications of our logic. First, we apply it to formally characterize different kinds of specifications of security properties. It has been observed that specifications are generally either *intensional*, i.e. based on details of the protocol steps, or *extensional*, i.e. formulated independently of message exchanges. We use the explicit belief modality to characterize intensional specifications, and the implicit belief modality to characterize extensional specifications.

Second, we show how to use these modalities to characterize and reason about attacks in interleaved protocol executions. Our specifications of security properties combine intensional and extensional specifications: the intensional part is used to represent the completion (or commitment) of agents in protocol executions, and the extensional part formalizes properties such as message secrecy. We illustrate this using the Needham-Schroeder Public Key protocol as a running example and show how the semantics can be used to demonstrate the existence of attacks. Afterwards we show the equivalence of our modal specification to those based on interleaved trace semantics.

We proceed following the structure given above. §2 gives the semantic foundations of our logic, and §3 and §4 discuss the two applications. We compare with related work in §5 (as explained there, this work supersedes our previous work on awareness-based security logics [2]), and conclude in §6.

2. A MULTI-MODAL SECURITY LOGIC

2.1 Syntax

We start by defining the set of messages, which are built from primitive terms by pairing and encryption. Based on this, we define a multi-modal language extended with operators expressing, e.g., possession and secrecy of messages.

Definition 1 Let the set $\mathcal{T}$ of *primitive terms* consist of three disjoint subsets: $\mathcal{T}_I$ of *agent identifiers*, $\mathcal{T}_K$ of *cryptographic keys*[1], and $\mathcal{T}_N$ of *nonces*. The set $\mathcal{M}$ of *messages* is the smallest set closed under the following rules: (i) $M \in \mathcal{M}$ if $M \in \mathcal{T}$; (ii) $M \circ M' \in \mathcal{M}$ if $M, M' \in \mathcal{M}$; and (iii) $\{\!|M|\!\}_K \in \mathcal{M}$ if $M \in \mathcal{M}$ and $K \in \mathcal{T}_K$.

The set $\mathcal{F}$ of *formulas* is the smallest set closed under the following rules: (i) $\bot \in \mathcal{F}$; (ii) $\varphi \to \psi \in \mathcal{F}$ if $\varphi \in \mathcal{F}$ and $\psi \in \mathcal{F}$; (iii) $\mathsf{says}_A(B, M)$, $\mathsf{sees}_A(M)$, $\mathsf{has}_A(M)$, $\mathsf{sec}_{\mathcal{G}}(M)$, $\mathsf{comml}_A(B, M), \mathsf{commR}_A(B, M) \in \mathcal{F}$ if $A, B \in \mathcal{T}_I$, $\mathcal{G} \subseteq \mathcal{T}_I$, and $M \in \mathcal{M}$; (iv) $\mathcal{X}_A\,\varphi$, $\mathcal{B}_A\,\varphi \in \mathcal{F}$ if $A \in \mathcal{T}_I$ and $\varphi \in \mathcal{F}$. ■

The formulas express properties of message exchanges between the agents engaged in a protocol execution (also called *run*). Intuitively, the formula $\mathsf{says}_A(B, M)$ denotes agent A *saying* M to B, $\mathsf{sees}_A(M)$ denotes A *seeing* M, $\mathsf{has}_A(M)$ denotes A *possessing* M, and $\mathsf{sec}_{\mathcal{G}}(M)$ denotes that M is a *secret* possessed only by the agents in the *group* $\mathcal{G}$. The formula $\mathsf{comml}_A(B, M)$ (respectively, $\mathsf{commR}_A(B, M)$) expresses that agent A uses message M to *commit* as the *initiator* (respectively, *responder*) in a protocol execution with agent B.[2] The modalities $\mathcal{X}_A$ and $\mathcal{B}_A$ denote the *explicit belief* and the *implicit belief* of an agent A. Other connectives and modalities are defined in the usual manner, e.g. negation $\neg\varphi \equiv \varphi \to \bot$ and conjunction $(\varphi \wedge \psi) \equiv \neg(\varphi \to \neg\psi)$.

To distinguish between the variables appearing in a protocol description and the actual values with which these variables are instantiated in a protocol execution, variables ranging over agent identifiers are denoted by capital letters $A, B, C, \ldots$, and the concrete values by lowercase letters $a, b, c, \ldots$, where the special constant *spy* denotes the *attacker*. We use the same convention also for keys and nonces, and write K and k, and N and n. We write $\mathcal{G}$ to denote a group of agents, α to denote an awareness set, and φ and ψ to denote formulas.

[1]We assume an underlying algebra where $(K^{-1})^{-1} = K$ for all keys $K \in \mathcal{T}_K$, and the function $\cdot^{-1} : \mathcal{T}_K \to \mathcal{T}_K$ maps a key K to its inverse key K^{-1}. For protocols employing (symmetric) shared keys we also have $K^{-1} = K$.

[2]The two predicates are used to express commitment of agents executing protocols with two roles, initiator and responder. It is straightforward to generalize the syntax and subsequent semantics with families of predicates like $\mathsf{comm}^k_j(A_1, \ldots, A_k, M)$, which formalizes commitment for the agent in the j-th role for a protocol with k roles.

2.2 Model of computation

Our model of computation combines ideas from trace-based methods for protocol verification [3, 18] with ideas from authentication logics [1, 6] and from approaches to reasoning about knowledge in multi-agent systems [10, 11, 22].

Trace-based foundations

An *event* e is a message exchange of the form $A \to B : M$ and a *trace* is a sequence $e_1, \ldots, e_k$ of *events* (where $\langle\rangle$ denotes the empty trace). A protocol is modeled as a set of traces; namely, the smallest set of traces closed under rules that formalize the effects of protocol steps and the possible actions by an attacker.

In Fig. 1 we show the NSPK protocol and its definition as an inductively defined set P of traces. The rules nspk_1, nspk_2, and nspk_3 formalize the three protocol steps. For example, nspk_2 models the second step of the protocol and says that a trace $t \in P$ can be extended with $B \to A : \{\!|N_A \circ N_B|\!\}_{K_A}$ whenever N_B has not been used in t (i.e. it is a *fresh* nonce) and t contains an event $A' \to B : \{\!|A \circ N_A|\!\}_{K_B}$. The $\mathsf{attacker}$ rule formalizes the attacker model of Dolev and Yao [9]: the *spy* can say anything that he can synthesize from the analyzable parts of the messages he spies, where $\mathsf{spies}(t)$ is the set consisting of all messages that have been sent in a trace t (which formalizes the assumption that the attacker has control over the network). The attacker rule uses the auxiliary functions synth and analz, which we now define (along with parts) as they will be needed in our model.

Definition 2 Let $\mathbf{M}$ be a set of messages. Using the corresponding rules in Fig. 2, we build the following three sets: the set $\mathsf{parts}(\mathbf{M})$ is the smallest extension of $\mathbf{M}$ obtained by adding the components of compound messages and the bodies of encrypted messages; the set $\mathsf{analz}(\mathbf{M})$ is the smallest extension of $\mathbf{M}$ closed under projection and decryption by keys in $\mathsf{analz}(\mathbf{M})$; and the set $\mathsf{synth}(\mathbf{M})$ is the smallest extension of $\mathbf{M}$ closed under pairing and encryption. ■

Modal foundations

The *local state* of an agent $A \in \mathcal{T}_I$ is a pair consisting of the set of actions that A has performed and the set of messages in A's possession. A *global state* w is an n-tuple of local states, where n is the number of agents in the system, including the attacker. In our model, the *actions* that an agent A can perform are *sending* a message M to another agent B, in symbols $\mathtt{send}_A(B, M)$, and *receiving* a message M, in symbols $\mathtt{rec}_A(M)$, where the identity of the sending agent is not known a priori.

We combine the notions of trace and state by defining functions that, given a trace, compute the local state of each agent participating in the (possibly partial, interleaved) protocol executions in the trace.

Definition 3 Given $t \in P$, the sets of actions and possessions of an agent A are defined by the functions $Ac_A(t)$ and

NSPK 1. $A \to B : \{|A \circ N_A|\}_{K_B}$

NSPK 2. $B \to A : \{|N_A \circ N_B|\}_{K_A}$

NSPK 3. $A \to B : \{|N_B|\}_{K_B}$

$$\frac{}{\langle\rangle \in P}\ \mathsf{empty} \qquad \frac{t \in P \quad N_A \notin \mathit{used}\ t}{t, A \to B : \{|A \circ N_A|\}_{K_B} \in P}\ \mathsf{nspk}_1$$

$$\frac{t \in P \quad N_B \notin \mathit{used}\ t \quad A' \to B : \{|A \circ N_A|\}_{K_B} \in t}{t, B \to A : \{|N_A \circ N_B|\}_{K_A} \in P}\ \mathsf{nspk}_2$$

$$\frac{t \in P \quad A \to B : \{|A \circ N_A|\}_{K_B} \in t \quad B' \to A : \{|N_A \circ N_B|\}_{K_A} \in t}{t, A \to B : \{|N_B|\}_{K_B} \in P}\ \mathsf{nspk}_3$$

$$\frac{t \in P \quad X \in \mathsf{synth}(\mathsf{analz}(\mathsf{spies}(t)))}{t, \mathit{spy} \to B : X \in P}\ \mathsf{attacker}$$

$ev_1 = a \to \mathit{spy} : \{|a \circ n_a|\}_{k_{spy}}$
$ev_2 = \mathit{spy} \to b : \{|a \circ n_a|\}_{k_b}$
$ev_3 = b \to a : \{|n_a \circ n_b|\}_{k_a}$
$ev_4 = a \to \mathit{spy} : \{|n_b|\}_{k_{spy}}$
$ev_5 = \mathit{spy} \to b : \{|n_b|\}_{k_b}$

Figure 1: The NSPK protocol (L), the rules defining it inductively (C), and the MITM attack on it (R)

$$\frac{M \in \mathbf{M}}{M \in \mathsf{parts}(\mathbf{M})}\ \mathsf{parts\text{-}inj} \qquad \frac{M_1 \circ M_2 \in \mathsf{parts}(\mathbf{M})}{M_i \in \mathsf{parts}(\mathbf{M})}\ \mathsf{parts}\text{-}i\ (i \in \{1,2\}) \qquad \frac{\{|M|\}_K \in \mathsf{parts}(\mathbf{M})}{M \in \mathsf{parts}(\mathbf{M})}\ \mathsf{parts\text{-}body}$$

$$\frac{M \in \mathbf{M}}{M \in \mathsf{analz}(\mathbf{M})}\ \mathsf{analz\text{-}inj} \qquad \frac{M_1 \circ M_2 \in \mathsf{analz}(\mathbf{M})}{M_i \in \mathsf{analz}(\mathbf{M})}\ \mathsf{analz}\text{-}i\ (i \in \{1,2\}) \qquad \frac{\{|M|\}_K \in \mathsf{analz}(\mathbf{M}) \quad K^{-1} \in \mathsf{analz}(\mathbf{M})}{M \in \mathsf{analz}(\mathbf{M})}\ \mathsf{analz\text{-}dec}$$

$$\frac{M \in \mathbf{M}}{M \in \mathsf{synth}(\mathbf{M})}\ \mathsf{synth\text{-}inj} \qquad \frac{M_1 \in \mathsf{synth}(\mathbf{M}) \quad M_2 \in \mathsf{synth}(\mathbf{M})}{M_1 \circ M_2 \in \mathsf{synth}(\mathbf{M})}\ \mathsf{synth\text{-}pair} \qquad \frac{M \in \mathsf{synth}(\mathbf{M}) \quad K \in \mathsf{synth}(\mathbf{M})}{\{|M|\}_K \in \mathsf{synth}(\mathbf{M})}\ \mathsf{synth\text{-}enc}$$

Figure 2: The rules defining the sets parts, analz and synth

$Po_A(t)$ as follows: $Ac_A(\langle\rangle) = \emptyset$ and $Ac_A(B \to C : M, ts)$ is

$$\begin{cases} \{\mathtt{send}_B(C, M)\} \cup Ac_A(ts) & \text{if } A = B \\ \{\mathtt{rec}_C(M)\} \cup Ac_A(ts) & \text{if } A = C \\ \{\mathtt{send}_B(C, M), \mathtt{rec}_C(M)\} \cup Ac_A(ts) & \text{if } A = \mathit{spy} \\ Ac_A(ts) & \text{otherwise} \end{cases}$$

and $Po_A(\langle\rangle) = \mathit{initState}(A)$ and $Po_A(B \to C : M, ts)$ is $\{M\} \cup Po_A(ts)$ if $A \in \{B, C, \mathit{spy}\}$ and $Po_A(ts)$ otherwise, where ts ranges over event sequences and *initState* is a protocol-dependent function returning the message items that an agent initially possesses (e.g. his private and public keys, and the public keys and identifiers of other agents).

Thus, given a trace $t \in P$, the local state $s_A(t)$ of an agent A is simply $\langle Ac_A(t), Po_A(t)\rangle$, and the global state w is the n-tuple of the local states $s_A(t)$ for all n agents. Given a global state w, we will (overloading notation) write $s_A(w)$ to denote the local state of an agent A at w, and $Ac_A(w)$ and $Po_A(w)$ to denote the two components of $s_A(w)$. ■

Hence, the *spy*'s local state contains the actions performed by all the agents, as well as the messages they exchange, while the local state $s_A(w)$ of an agent A different from the *spy* is built only from the events that A participated in. Since the *spy* possesses all the messages sent in the network, $Po_{spy}(w)$ captures the same information as the set spies used in Fig. 1 to formalize the attacker's control over the network.

Let t be a trace and ts be the sequence of all prefixes of t. The set W^t of global states (or *worlds*) relative to t is obtained by computing, for each prefix of t, the corresponding sets of actions and possessions for all agents A. Formally, $W^t = \mathit{wrl}(ts)$, where

$$\mathit{wrl}(ts) = \begin{cases} \{(Ac_A(t'), Po_A(t'))\} \cup \mathit{wrl}(ts') & \text{if } ts = t', ts' \\ \{(Ac_A(\langle\rangle), Po_A(\langle\rangle))\} & \text{if } ts = \langle\rangle \end{cases}$$

Modeling resource-bounded agents

In the artificial intelligence literature, resource-bounded agents have limited computational resources, such as memory or time. In our approach, we model resource-bounded agents where the limitations are both in (1) the propositions an agent may reason about (his language) and (2) his deductive ability (what he can conclude). As an example of (1), if a nonce N is a secret between A and B, then another agent C should not even be able to formulate propositions about it. As an example of (2), when C learns the nonce N, he can then conclude that he possesses it, but not necessarily that some other agent D possesses it (even when this is the case).

Our first step in limiting resources is to restrict an agent's language by making the messages he can construct a function of the information he possesses at a state.

Definition 4 The set $\mathcal{M}_A(w)$ of messages that an agent A can construct at a global state w is defined as $\mathcal{M}_A(w) = \{M \mid M \in \mathsf{synth}(\mathsf{analz}(Po_A(w)))\}$. The set $\mathcal{F}_A(w)$ of formulas of an agent A at a global state w is the smallest set of formulas closed under the following rules: (i) $\bot \in \mathcal{F}_A(w)$; (ii) $\varphi \to \psi \in \mathcal{F}_A(w)$ if $\varphi, \psi \in \mathcal{F}_A(w)$; (iii) $\mathsf{says}_B(C, M)$, $\mathsf{sees}_B(M)$, $\mathsf{has}_B(M)$, $\mathsf{sec}_{\mathcal{G}}(M)$, $\mathsf{comml}_B(C, M)$, $\mathsf{commR}_B(C, M) \in \mathcal{F}_A(w)$ if $B, C \in \mathcal{M}_A(w) \cap \mathcal{T}_I$, $M \in \mathcal{M}_A(w)$, and $\mathcal{G} \subseteq \mathcal{M}_A(w) \cap \mathcal{T}_I$; and (iv) $\mathcal{X}_A\varphi \in \mathcal{F}_A(w)$ if $\varphi \in \mathcal{F}_A(w)$. ■

Clause (iii) expresses that each agent has its own language for the predicates says, sees, has, sec, comml, and commR, which depends on the messages that an agent possesses at some state w. In comparison with rule (iii) in Def. 1, here we simply require that the message items belong to the set of messages of the agent. For example, $\mathsf{says}_A(B, M) \in \mathcal{F}_A(w)$ if A and B are agent identifiers in $\mathcal{M}_A(w)$ (denoted by

$A, B \in \mathcal{M}_A(w) \cap \mathcal{T}_I$) and M is a message in the set of messages of A (in symbols $M \in \mathcal{M}_A(w)$).

With respect to (iv), note that since the agents' languages do not include the modality $\mathcal{B}$ for implicit belief, an agent can reason about neither his own nor other agents' implicit beliefs, nor can he have explicit beliefs about the explicit beliefs of other agents (as is standard in belief logics, e.g. [15]).

2.3 Semantics

We begin by fixing a set $\overline{\mathcal{T}}_I$ of *agent names*, where, for notational simplicity, we identify its elements with the previously defined set $\mathcal{T}_I$ of agent identifiers; thus, from now on we will simply talk of agents. Similarly, for keys and nonces.

Given a trace $t \in P$, we obtain the corresponding model $\mathfrak{M}^t = (W^t, \sim, \alpha)$, where W^t is a non-empty set of worlds, $\sim$ is an agent-indexed family of equivalence relations on W^t, and α is an agent-indexed family of awareness sets, where the set $\alpha_A(w)$ consists of the formulas that agent A is aware of at world w. The family of equivalence relations $\sim$ captures *indistinguishability*: two global states are indistinguishable to an agent A iff the local state of A is the same at these two global states. Formally, $w \sim_A w'$ iff $s_A(w) = s_A(w')$, i.e. $Ac_A(w) = Ac_A(w')$ and $Po_A(w) = Po_A(w')$. Note that our model does not contain a valuation function as we do not have propositional symbols.

A protocol execution results from agents taking actions and corresponds to a multi-agent system. We can view this system from two perspectives: that of an external agent who observes the system from the outside and does not interact with the agents executing the protocol, and that of the internal agents engaged in the execution. The former view is formalized using a *global truth relation*, denoted by $\models$. The latter is formalized by a *local truth relation*, which is a family of truth relations $\models_A$, indexed by agents A.

2.3.1 Global truth

The global truth relation formalizes what an external observer can conclude from the system. By design, this agent is not resource-bounded and has access to all communication and can reason about the local states of individual agents. In particular, he ascribes implicit belief to the agents, i.e. he can compute whether an agent A would implicitly believe in some formula φ, had A enough information about the overall communication that is taking place.

In order to formalize these ideas, and to define the semantics for predicates such as says and sees, we need to express specific relationships between agents and messages at a global state. For example, an agent should only be entitled to say the messages he is able to compose from the information he possesses. Similarly, he should be entitled to see the submessages that he can obtain from a message he receives. To this end, we introduce the operators comp and submsg to define two abbreviations that will be useful in the semantic definitions in §2.3; assuming that M is a message at w, the set $\mathsf{comp}_A(w, M)$ contains all the sub-messages that A used to construct the message M at w, i.e. $\mathsf{comp}_A(w, M) =$

$$\begin{cases} Po_A(w) \cap \mathsf{parts}(\{M\}) & \text{if } M \in \mathcal{M}_A(w) \\ \emptyset & \text{otherwise} \end{cases},$$

and the set $\mathsf{submsg}_A(w, M)$ consists of all sub-messages that A can obtain from M given the keys he possesses at w, i.e.

$$\mathsf{submsg}_A(w, M) = \mathsf{analz}(Po_A(w)) \cap \mathsf{parts}(\{M\})\,.$$

We use *commit sets* $\mathcal{C}$ to define the semantics of the commI and commR formulas. During one execution of a protocol *Prot*, an agent A can take either the initiator role or the responder role. Intuitively, within an execution, each role is identified by some message M, where M is, or contains, a nonce. The set $\mathcal{C}_I^{Prot}(A, B, M)$ contains the actions that A performed in order to commit as initiator to a responder B using message M. Similarly, $\mathcal{C}_R^{Prot}(A, B, M)$ contains the actions that B performed in order to commit as responder to an initiator A using message M. Both sets are obtained directly from the description of the protocol. We illustrate this by means of our running example.

Example 5 In an execution of the NSPK protocol (Fig. 1), the initiator A commits to the responder B using N_A after performing the actions corresponding to the steps encoded by the rules nspk_1, nspk_2 and nspk_3. Hence, the commit set $\mathcal{C}_I^{\text{NSPK}}(A, B, N_A) = \{\mathtt{send}_A(B, \{\!|A \circ N_A|\!\}_{K_B}), \mathtt{rec}_A(\{\!|N_A \circ N_B|\!\}_{K_A}), \mathtt{send}_A(B, \{\!|N_B|\!\}_{K_B})\}$. Similarly, the responder's "view" is formalized by the set $\mathcal{C}_R^{\text{NSPK}}(A, B, N_B) = \{\mathtt{rec}_B(\{\!|A \circ N_A|\!\}_{K_B}), \mathtt{send}_B(A, \{\!|N_A \circ N_B|\!\}_{K_A}), \mathtt{rec}_B(\{\!|N_B|\!\}_{K_B})\}$. ■

We are now ready to define the global truth relation.

Definition 6 The truth of a formula φ at a global state w in a model $\mathfrak{M} = (W, \sim, \alpha)$, in symbols $\mathfrak{M}, w \models \varphi$, is the smallest relation satisfying:

$$\begin{array}{rll}
\mathfrak{M}, w \models \mathsf{says}_A(B, M) & \text{if} & \mathtt{send}_A(B, M') \in Ac_A(w) \text{ and } M \in \mathsf{comp}_A(w, M') \text{ for some } M' \\
\mathfrak{M}, w \models \mathsf{sees}_A(M) & \text{if} & \mathtt{rec}_A(M') \in Ac_A(w) \text{ and } M \in \mathsf{submsg}_A(w, M') \text{ for some } M' \\
\mathfrak{M}, w \models \mathsf{has}_A(M) & \text{if} & M \in \mathsf{analz}(Po_A(w)) \\
\mathfrak{M}, w \models \mathsf{sec}_{\mathcal{G}}(M) & \text{if} & \mathfrak{M}, w \models \mathsf{has}_A(M) \text{ for all } A \in \mathcal{G} \text{ and } \mathfrak{M}, w \not\models \mathsf{has}_B(M) \text{ for all } B \notin \mathcal{G} \\
\mathfrak{M}, w \models \mathsf{commI}_A(B, M) & \text{if} & \mathcal{C}_I^{Prot}(A, B, M) \subseteq Ac_A(w) \\
\mathfrak{M}, w \models \mathsf{commR}_A(B, M) & \text{if} & \mathcal{C}_R^{Prot}(B, A, M) \subseteq Ac_A(w) \\
\mathfrak{M}, w \models \varphi \to \psi & \text{if} & \mathfrak{M}, w \not\models \varphi \text{ or } \mathfrak{M}, w \models \psi \\
\mathfrak{M}, w \models \mathcal{B}_A \varphi & \text{if} & \mathfrak{M}, w' \models \varphi \text{ for all } w' \text{ such that } w \sim_A w' \\
\mathfrak{M}, w \models \mathcal{X}_A \varphi & \text{if} & \mathfrak{M}, w \models_A \varphi \text{ and } \varphi \in \mathcal{F}_A(w)
\end{array}$$

We write $\mathfrak{M} \models \varphi$ iff $\mathfrak{M}, w \models \varphi$ for all $w \in W$. ■

In other words, at a global state w an agent A *says* M to an agent B iff he sent an M' to B such that he used M in composing M', A *sees* M iff he received an M' such that M is a readable sub-message of M', and A *has* M iff M is an analyzable message in A's set of possessions. A message M is a *secret* shared among the agents in a group $\mathcal{G}$ at w iff at w all the agents in $\mathcal{G}$ possess M and no agent outside the group possesses M. Moreover, A *commits* to an agent B as an initiator (respectively, responder) using M iff A has performed the actions in the initiator's (respectively, responder's) commit set. Furthermore, an agent A *implicitly believes* in φ at

w iff φ holds in all the worlds indistinguishable to A from w, which is the standard interpretation of the belief of logically omniscient agents. We employ the explicit belief modality (and awareness) to formalize the formulas in which a non-omniscient, resource-bounded agent believes: We start by restricting the formulas φ that an agent might explicitly believe in at a global state w to those in his language, which is expressed by $\varphi \in \mathcal{F}_A(w)$ (see Def. 4), and then further restrict these formulas to those he can prove using the information he currently possesses, which is captured using the local truth relation $\models_A$.

2.3.2 Local truth

$\mathfrak{M}, w \models_A \varphi$ captures the truth of a formula φ relative to an agent A at a global state w. Since there are situations in which φ expresses properties of A himself, and situations in which φ expresses properties of other agents, we will distinguish between these two forms of reasoning in our definition below. In particular, different forms of reasoning require different kinds of information. For example, if A has to check whether he possesses M, he will check whether his possession set contains M. But, to check whether an agent B has M, A cannot just access the set of B's possessions. In our formalization, A uses his awareness set to determine whether B used M to compose a message B has sent, or that B received M in some message M' that B can analyze.

Modeling an agent reasoning about his own local state is straightforward: we use the sets comp, submsg and analz to define the semantics for the says, sees and has predicates, respectively, as in Def. 6.

Modeling an agent reasoning about other agents is more complicated. Here we employ the agent's awareness set to define the semantics of the formulas. To accomplish this, we define "meta-versions" of the sets comp and submsg, expressing the messages that some other agent may have used to compose a message he has sent, as well as the sub-messages he might be able to obtain from a message he has received. These capabilities are formalized by means of the sets m-comp and m-submsg, respectively. The set $\mathsf{m\text{-}comp}_A(B, C, w, M)$ consists of the messages that, at global state w, A expects B to have used to send the message M to some agent C. The set $\mathsf{m\text{-}submsg}_A(B, w, M)$ consists of the sub-messages of M that, according to A's awareness set at w, agent B might be able to possess. The rules defining these sets are given in Fig. 3.

We explain the intuition behind some these rules. Rule mc-inj formalizes that if an agent A is aware that an agent B sent a message M, then M is among the messages that A expects B to have sent. In ms-pk, if A observed that B received a message M' such that M encrypted with B's public key K_B is part of M', then A concludes that B has M. Note that, although an agent reasons about messages that he may be unable to analyze, there will not be any secrecy violation following from these rules: reasoning about the existence of a message does not correspond to possessing it.

The awareness set of an agent encodes the actions that he expects other agents to have performed. To reason about commitment, we have to check whether a set of actions, i.e. a commit set, is a subset of the awareness set of an agent. To this end, we introduce a function that maps actions ac in a set $\mathcal{C}$ to the corresponding set $form(\mathcal{C}) = \{a2f(ac) \mid ac \in \mathcal{C}\}$ of formulas, where $a2f(ac) = \mathsf{says}_A(B, M)$ if $ac = \mathtt{send}_A(B, M)$ and $a2f(ac) = \mathsf{sees}_A(M)$ if $ac = \mathtt{rec}_A(M)$.

We now turn to the formal definition of $\models_A$.

Definition 7 The truth of a formula φ relative to an agent A at a global state w in a model $\mathfrak{M} = (W, \sim, \alpha)$, in symbols $\mathfrak{M}, w \models_A \varphi$, is the smallest relation satisfying:

$$\mathfrak{M}, w \models_A \varphi \to \psi \quad \text{if} \quad \mathfrak{M}, w \not\models_A \varphi \text{ or } \mathfrak{M}, w \models_A \psi$$

For an agent reasoning about himself:[3]

$$\begin{aligned}
\mathfrak{M}, w \models_A \mathsf{says}_A(B, M) \quad &\text{if} \quad \mathtt{send}_A(B, M') \in Ac_A(w) \\
&\text{and } M \in \mathsf{comp}_A(w, M') \text{ for some } M' \\
\mathfrak{M}, w \models_A \mathsf{sees}_A(M) \quad &\text{if} \quad \mathtt{rec}_A(M') \in Ac_A(w) \\
&\text{and } M \in \mathsf{submsg}_A(w, M') \text{ for some } M' \\
\mathfrak{M}, w \models_A \mathsf{has}_A(M) \quad &\text{if} \quad M \in \mathsf{analz}(Po_A(w)) \\
\mathfrak{M}, w \models_A \mathsf{comml}_A(B, M) \quad &\text{if} \quad \mathcal{C}_I^{Prot}(A, B, M) \subseteq Ac_A(w) \\
\mathfrak{M}, w \models_A \mathsf{commR}_A(B, M) \quad &\text{if} \quad \mathcal{C}_R^{Prot}(B, A, M) \subseteq Ac_A(w)
\end{aligned}$$

For an agent A reasoning about an agent $B \neq A$:

$$\begin{aligned}
\mathfrak{M}, w \models_A \mathsf{says}_B(C, M) \quad &\text{if} \quad \mathsf{says}_B(C, M') \in \alpha_A(w) \\
&\text{and } M' \text{ such that } M \in \mathsf{m\text{-}comp}_A(B, C, w, M') \\
\mathfrak{M}, w \models_A \mathsf{sees}_B(M) \quad &\text{if} \quad \mathsf{sees}_B(M') \in \alpha_A(w) \\
&\text{and } M' \text{ such that } M \in \mathsf{m\text{-}submsg}_A(B, w, M') \\
\mathfrak{M}, w \models_A \mathsf{has}_B(M) \quad &\text{if} \quad \mathfrak{M}, w \models_A \mathsf{says}_B(C, M) \text{ for} \\
&\text{some } C \text{ or } \mathfrak{M}, w \models_A \mathsf{sees}_B(M) \\
\mathfrak{M}, w \models_A \mathsf{comml}_B(C, M) \quad &\text{if} \quad form(\mathcal{C}_I^{Prot}(B, C, M)) \subseteq \alpha_A(w) \\
&\text{for some } C \\
\mathfrak{M}, w \models_A \mathsf{commR}_B(C, M) \quad &\text{if} \quad form(\mathcal{C}_R^{Prot}(C, B, M)) \subseteq \alpha_A(w) \\
&\text{for some } C
\end{aligned}$$

The semantics for secrecy (where the agent identifiers range over the identifiers in A's possession set $Po_A(w)$) is:

$$\mathfrak{M}, w \models_A \mathsf{sec}_{\mathcal{G}}(M) \quad \text{if} \quad \mathfrak{M}, w \models_A \mathsf{has}_B(M) \text{ for all } B \in \mathcal{G} \text{ and } \mathfrak{M}, w \not\models_A \mathsf{has}_C(M) \text{ for all } C \notin \mathcal{G}. \quad \blacksquare$$

Let us give the intuition behind the definitions for an agent A reasoning about another agent B. We define that, for an agent A at global state w, B *says* M to an agent C iff A is aware that B has sent an M' to C such that M was (expected to be) used by B to compose M'. Similarly, agent B *sees* a message M iff A is aware that B has received a message M' such that M is a sub-message that B is (expected to be) able to see from M'. Agent B *has* M iff either B says or sees M. From the point of view of A, an agent B has committed to an agent C as an initiator of an execution identified by M iff A is aware that B has performed the actions in the initiator's commit set. Similarly, for the commR formula. As we observed above, there is no clause for explicit belief since an agent cannot reason about what another agent may explicitly believe.

Note that an agent reasoning about his own state (local truth) coincides with an external agent reasoning about this

[3]We do not define the $\models_A$ relation in the case of the $\mathcal{X}_A$ since this reduces trivially to $\models_A$.

$$\frac{\mathsf{says}_B(C,M) \in \alpha_A(w)}{M \in \mathsf{m\text{-}comp}_A(B,C,w,M)}\ \text{mc-inj} \qquad \frac{\{\!|M|\!\}_{K_B^{-1}} \in \mathsf{m\text{-}comp}_A(B,C,w,M')}{M \in \mathsf{m\text{-}comp}_A(B,C,w,M')}\ \text{mc-sig} \qquad \frac{M_1 \circ M_2 \in \mathsf{m\text{-}comp}_A(B,C,w,M)}{M_i \in \mathsf{m\text{-}comp}_A(B,C,w,M)}\ \text{mc-}i\ (i \in \{1,2\})$$

$$\frac{\mathsf{sees}_B(M) \in \alpha_A(w)}{M \in \mathsf{m\text{-}submsg}_A(B,w,M)}\ \text{ms-inj} \qquad \frac{\{\!|M|\!\}_{K_B} \in \mathsf{m\text{-}submsg}_A(B,w,M')}{M \in \mathsf{m\text{-}submsg}_A(B,w,M')}\ \text{ms-pk} \qquad \frac{M_1 \circ M_2 \in \mathsf{m\text{-}submsg}_A(B,w,M)}{M_i \in \mathsf{m\text{-}submsg}_A(B,w,M)}\ \text{ms-}i\ (i \in \{1,2\})$$

Figure 3: The rules defining the sets m-comp and m-submsg

agent (global truth). Hence, as shown in the appendix, it follows straightforwardly from Def. 6 and Def. 7 that:

Lemma 8 For all agents A and B, global states w, and formulas $\varphi \in \mathcal{F}_A(w)$ such that $\varphi \in \{\mathsf{says}_A(B,M), \mathsf{sees}_A(M), \mathsf{has}_A(M)\}$, we have that $\mathfrak{M}, w \models \mathcal{B}_A\,\varphi$ iff $\mathfrak{M}, w \models \mathcal{X}_A\,\varphi$. ■

To summarize, our formalization expresses that there are two sources of information (local states and awareness sets), which provide different levels of reliability (certainties and expectations) and are employed differently (for reasoning about oneself or about other agents).

2.4 Defining awareness

We use awareness to represent the expectations of an agent with respect to the actions of the agents with whom he is communicating. Each step of a protocol gives rise to (i) a rule capturing the expectations of the sender with respect to the `send` action he has performed, and (ii) a rule capturing the expectations of the receiver regarding the corresponding `rec` action. Note that an agent's expectations may not correspond to reality, as he might be aware of, and thus explicitly believe in, false statements (as is the case in the man-in-the-middle attack on the NSPK protocol, which we consider below).

The rules representing the sender perspective are obtained from the protocol steps in a straightforward manner. Given the n-th step $A \to B : M$ of a protocol *Prot*, the sender A, who has the $\mathtt{send}_A(B,M)$ action recorded in his local state, expects the receiver B to get the message M; thus, the rule $Prot\_s_n$ adds the formula $\mathsf{sees}_B(M)$ to A's awareness set:

$$\frac{\mathtt{send}_A(B,M) \in Ac_A(w)}{\mathsf{sees}_B(M) \in \alpha_A(w)}\ Prot\_s_n\,.$$

The rules capturing the expectations of the receiver depend on the protocol the agents are executing, and thus cannot be given in a general form like the sender rule. Instead, we consider a concrete example and give the receiver rules for the NSPK protocol in Fig. 4.

The intuition behind the rule nspk_r_1 is that, upon the receipt of the first message, agent B expects that it has been sent by agent A. The rule nspk_r_2 formalizes that when A receives his nonce N_A back, he may conclude that B sent it. The intuition behind nspk_r_3 is similar.

Note that the expectations of the agents do not always correspond to what is actually happening. In fact, attacks run counter to the expectations of the agents (as these are based on incomplete information).

Although illustrated only for the NSPK protocol, the ideas presented here are general. We have used our logic to reason about a number of other protocols, e.g. the full NSPK protocol and the Otway-Rees protocol with shared keys.

3. MODALITIES AND SPECIFICATIONS

In this section, we use our modalities to formally characterize different kinds of specifications of security properties. Furthermore, we show how to use them to reason semantically about attacks in interleaved protocol executions.

3.1 Extensional and intensional specifications

A number of researchers, e.g. [5, 12, 19], have observed that there are two different kinds of security specifications: extensional specifications, which are, in some sense, independent of the details of a particular protocol, and intensional specifications, where statements of properties are based on the protocol itself. For example, consider the definitions given by Roscoe [19, pages 31 and 34]:

> We classify a specification as *extensional* when it is independent of the details of the protocol and would apply to any other protocol designed to achieve the same effect. Thus, inevitably, it cannot mention the actual messages passing between nodes during a protocol since these vary from one to another. Instead, it will test the states of mind (knowledge, belief, etc.) of the various participants including the spy.
>
> A specification is classified as *intensional* when its primary purpose is to assert a property of the way, in terms of communications within a protocol, a particular state is reached.

Until now, these definitions have lacked a formal status. One of the contributions of our work is to characterize these notions in terms of our modalities.

We begin by observing that the implicit belief modality has an extensional character as the properties it formalizes are independent of the particular message exchanges. Intuitively, this is because implicit belief captures the view of an external, resource-unbounded observer following the protocol execution. In order to check whether a property denoted by a formula φ holds, such an observer need not be aware of the particular message exchanges of the protocol execution; rather, he simply checks whether the local states of the agents satisfy φ.

$$\frac{\mathtt{rec}_B(\{\!|A \circ N_A|\!\}_{K_B}) \in Ac_B(w)}{\mathsf{says}_A(B, \{\!|A \circ N_A|\!\}_{K_B}) \in \alpha_B(w)}\ \mathsf{nspk\_}r_1 \qquad \frac{\begin{array}{c}\mathtt{send}_A(B, \{\!|A \circ N_A|\!\}_{K_B}) \in Ac_A(w) \\ \mathtt{rec}_A(\{\!|N_A \circ N_B|\!\}_{K_A}) \in Ac_A(w)\end{array}}{\mathsf{says}_B(A, \{\!|N_A \circ N_B|\!\}_{K_A}) \in \alpha_A(w)}\ \mathsf{nspk\_}r_2 \qquad \frac{\begin{array}{c}\mathtt{send}_B(A, \{\!|N_A \circ N_B|\!\}_{K_A}) \in Ac_B(w) \\ \mathtt{rec}_B(\{\!|N_B|\!\}_{K_B}) \in Ac_B(w)\end{array}}{\mathsf{says}_A(B, \{\!|N_B|\!\}_{K_B}) \in \alpha_B(w)}\ \mathsf{nspk\_}r_3$$

Figure 4: Receiver rules for the NSPK protocol

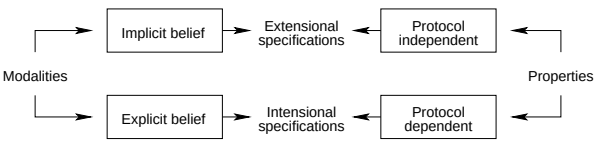

Figure 5: Relation between modalities, kinds of specification and security properties

In contrast, the explicit belief modality has an intensional character. This modality is based on the agents' awareness sets, which model the agents' local, resource-bounded views of the expected results of their actions. The awareness sets are in turn determined by the protocol rules, and hence statements about explicit belief are statements about the results of particular protocol steps. Fig. 5 summarizes these relationships.

This logical characterization of the two definitions has the status of a thesis: since the definitions are informal (natural language), our thesis cannot be formally proven. However, we can support it by showing that it holds for different commonly considered kinds of specifications. In what follows, we will consider two examples: *secrecy* of a message (which can be specified extensionally) and an agent's *completion* of a protocol execution (which is inherently intensional).

To illustrate how these properties can be modally specified, we return to the NSPK protocol, our running example. We will start from an informal specification of a correctness requirement for a responder of the protocol, expressed in terms of secrecy and completion. We then show how a progressive refinement of the informal specifications of these properties leads to their formal specifications in terms of implicit and explicit belief, supporting our thesis. Moreover, in §3.3 and §4, we will use the resulting specification to show how we can reason about modal specifications, and in what sense this specification is equivalent to one stated directly in terms of properties of interleaved traces.

3.2 Formalizing secrecy and completion

A message is a secret between a group of agents at some state of a protocol execution when, at that state, all the agents in the group have the message, and all the remaining agents do not have it. The way in which the state is reached is irrelevant to the secrecy of a message. Only the possession of the agents at that particular state matters! On the other hand, the completion of a protocol execution by an agent refers to a sequence of actions performed by the agent, not to the properties of an individual state. Thus, a specification of completion involves the details of the protocol and is therefore intensional.

Consider again the NSPK protocol. Agent B uses his nonce N_B as a challenge to authenticate agent A. One way of expressing the correctness requirement for a responder B executing an instance of the NSPK protocol with an initiator A is by means of the following two properties that must hold in a trace t:

P1. *Completion*: B completes an execution with A using the nonce N_B.

P2. *Secrecy*: The responder's nonce N_B is a secret between B and the initiator A.

Note that secrecy is necessary, but not sufficient, for correctness; we should check whether the responder's nonce N_B is a secret only when the corresponding protocol execution is completed. That is, property 1 should imply property 2.

We refine these informal specifications and formalize them within our logic. We start with the completion property, which we state in a more refined but still informal way.

P1'. There exists a state in the trace such that B completes an execution as a responder with an initiator A using a nonce N_B.

Formalizing the meaning of "B completes an execution with an initiator A using a nonce N_B", we obtain:

P1". There exists a state w of $\mathfrak{M}^t$ such that $\mathfrak{M}^t, w \models \mathcal{X}_B\, \mathsf{commR}_B(A, N_B)$.

More specifically, the predicate commR formalizes that B has completed the execution as a responder, and the intensionality is represented by the use of the explicit belief modality.

We now turn to the secrecy property, which (bringing out its inherent extensionality) we can state in a more refined but still informal way as:

P2'. There exists a state in the trace such that N_B is a secret between agents B and A.

The formalization of this extensional specification is based on the fact that, in order to specify what a secret is, we do not need to refer to the protocol: N_B is a secret between B and A iff B and A are the only agents who possess N_B. We can directly formalize this using the sec predicate and the implicit belief modality to obtain:

P2". There exists a state w of $\mathfrak{M}^t$ such that $\mathfrak{M}^t, w \models \mathcal{B}_B \, \mathsf{sec}_{\{A,B\}}(N_B)$.

The correctness requirement for the responder of an execution of the NSPK protocol combines completion and secrecy by requiring that if, for some trace t, B completes an execution identified by N_B as a responder of A, then N_B must be a secret between B and A. Formally, the responder's requirement for the protocol is: for all $t \in P$, models $\mathfrak{M}^t$, agents A and B, and nonces N_B,

$$\mathfrak{M}^t \models \mathcal{X}_B \, \mathsf{commR}_B(A, N_B) \rightarrow \mathcal{B}_B \, \mathsf{sec}_{\{A,B\}}(N_B) \, . \qquad (1)$$

3.3 Reasoning about the NSPK protocol

We now show how to use our semantics to reason about specifications. There are two possibilities, depending on the relationship of the specification to the set of intended models (i.e. traces). The first possibility is verification, which is establishing the correctness of a protocol by showing that it holds for all "protocol conform" models. For instance, we could show that all models $\mathfrak{M}^t$, resulting from all possible NSPK protocol traces t, satisfy the agents' requirements, such as that for B in (1). In a manner similar to Paulson's inductive method, such verification could be carried out by induction over the set of all models $\mathfrak{M}^t$, corresponding to those t in the inductively defined set of traces.

A second possibility is falsification. We will illustrate this here by giving a model where B's requirement (1) fails to hold. That is, to falsify (1), we give a particular model $\mathfrak{M}^t_{\text{MITM}}$ that models an execution trace corresponding to the man-in-the-middle (MITM) attack.

Theorem 9 There exist an NSPK execution trace t, a model $\mathfrak{M}^t$, agents A and B, and a nonce N_B such that $\mathfrak{M}^t \not\models \mathcal{X}_B \, \mathsf{commR}_B(A, N_B) \rightarrow \mathcal{B}_B \, \mathsf{sec}_{\{A,B\}}(N_B)$. ■

The MITM attack on the NSPK protocol [16] consists of the sequence of events shown in Fig. 1. Thus, consider the model $\mathfrak{M}^t_{\text{MITM}} = (W^t, \sim, \alpha)$ obtained from the trace $t = \langle ev_1, ev_2, ev_3, ev_4, ev_5 \rangle$, which represents the smallest sequence of events containing this attack. The components of $\mathfrak{M}^t_{\text{MITM}}$ that are relevant for our analysis are:

- $W^t = \{w_0, w_1, w_2, w_3, w_4, w_5\}$,
- $\sim_a = \{(w_0,w_0),(w_1,w_1),(w_1,w_2),(w_2,w_1),(w_2,w_2),$ $(w_3,w_3),(w_4,w_4),(w_4,w_5),(w_5,w_4),(w_5,w_5)\}$,
- $\sim_b = \{(w_0,w_0),(w_0,w_1),(w_1,w_0),(w_1,w_1),(w_2,w_2),$ $(w_3,w_3),(w_3,w_4),(w_4,w_3),(w_4,w_4),(w_5,w_5)\}$,
- $\alpha_a(w_5) = \{\mathsf{sees}_{spy}(\{\!|a \circ n_a|\!\}_{k_{spy}}), \mathsf{says}_{spy}(a, \{\!|n_a \circ n_b|\!\}_{k_{spy}}),$ $\mathsf{sees}_{spy}(\{\!|n_b|\!\}_{k_{spy}})\}$,
- $\alpha_b(w_5) = \{\mathsf{says}_a(b, \{\!|a \circ n_a|\!\}_{k_b}), \mathsf{sees}_a(\{\!|n_a \circ n_b|\!\}_{k_a}),$ $\mathsf{says}_a(b, \{\!|n_b|\!\}_{k_b})\}$.

We focus on w_5 since it is the global state obtained after the last event in t. The local states of the agents a and b and the possessions of the *spy* at w_5 are:

$$s_a(w_5) = \begin{cases} Po_a(w_5) = \{a, k_a, k_a^{-1}, spy, k_{spy}, n_a, n_b\}, \\ Ac_a(w_5) = \{\mathtt{send}_a(spy, \{\!|a \circ n_a|\!\}_{k_{spy}}), \\ \quad \mathtt{rec}_a(\{\!|n_a \circ n_b|\!\}_{k_a}), \mathtt{send}_a(spy, \{\!|n_b|\!\}_{k_{spy}})\} \end{cases},$$

$$s_b(w_5) = \begin{cases} Po_b(w_5) = \{b, k_b, k_b^{-1}, a, k_a, n_a, n_b\} \\ Ac_b(w_5) = \{\mathtt{rec}_b(\{\!|a \circ n_a|\!\}_{k_b}), \\ \quad \mathtt{send}_b(a, \{\!|n_a \circ n_b|\!\}_{k_a}), \mathtt{rec}_b(\{\!|n_b|\!\}_{k_b})\} \end{cases},$$

$$Po_{spy}(w_5) = \{spy, k_{spy}, k_{spy}^{-1}, a, k_a, b, k_b, n_a, n_b\},$$

where $initState(a) = \{a, k_a, k_a^{-1}, spy, k_{spy}\}$, $initState(b) = \{k_a, b, k_b, k_b^{-1}\}$ and $initState(spy) = \{spy, k_{spy}, k_{spy}^{-1}, a, k_a, b, k_b\}$. We can then use the semantics to demonstrate the existence of the attack (as shown in the proof of Theorem 9 in the appendix).

4. MODAL VERSUS TRACE-BASED SPECIFICATIONS

We now show that our modal specification for B's correctness requirement for the NSPK protocol is equivalent to a trace-based specification of the same protocol requirement. We establish this by showing the logical equivalence of both specifications with respect to our semantics.

As noted above, a trace-based interleaved semantics can be used both for interactive verification [18] and for falsification based on infinite-state model-checking [3]. The specifications in both approaches are intensional and specify what must (or cannot) hold after certain occurrences of events. For example, for verification, in the case of NSPK one might specify B's requirement by formalizing that N_B is a secret after the last two steps of the protocol have occurred:

$$(\mathsf{sees}_B(\{\!|N_B|\!\}_{K_B}) \wedge \mathsf{says}_B(A, \{\!|N_A \circ N_B|\!\}_{K_A})) \rightarrow \neg \mathsf{has}_{spy}(N_B) \, . \qquad (2)$$

For falsification, one formalizes the negation of (2), i.e.

$$(\mathsf{sees}_B(\{\!|N_B|\!\}_{K_B}) \wedge \mathsf{says}_B(A, \{\!|N_A \circ N_B|\!\}_{K_A})) \wedge \mathsf{has}_{spy}(N_B) \, , \qquad (3)$$

and searches for a trace with this property.

The specification (3) is a direct translation of the Haskell program used in [3] to specify an attack.[4] This can be directly expressed as a formula in our logic and proved for some A, B, and N_B, at some world w of some model $\mathfrak{M}^t$ resulting from some execution trace t.

Showing that this is equivalent to the statement of Theorem 9 establishes that the attack in the trace-based specification is equivalent to the attack in our modal specification with respect to our semantics. The equivalence between the two specifications can be shown alternatively (in terms of "verification" rather than "falsification") by showing the

[4]Paulson's verification specification is similar to (2). He formalizes an intensional specification of secrecy for the nonce N_B by stating that if there is an event $B \rightarrow A : \{\!|N_A \circ N_B|\!\}_{K_A}$ in the set of traces modeling the NSPK protocol, then the *spy* does not possess the nonce N_B.

equivalence of (1) and of a formula representing the correctness of B's requirement. As shown in the appendix, for non-compromised agents we have:

Theorem 10 For all traces t of the NSPK protocol, models $\mathfrak{M}^t$, agents A, B such that $A \neq spy$, $B \neq spy$, $\mathfrak{M}^t \models \neg\mathsf{has}_{spy}(K_A^{-1})$ and $\mathfrak{M}^t \models \neg\mathsf{has}_{spy}(K_B^{-1})$, and nonces N_B,

$$\mathfrak{M}^t \models \mathcal{X}_B\,\mathsf{commR}_B(A, N_B) \rightarrow \mathcal{B}_B\,\mathsf{sec}_{\{A,B\}}(N_B)$$

iff

$$\mathfrak{M}^t \models (\mathsf{sees}_B(\{\!|N_B|\!\}_{K_B}) \wedge \mathsf{says}_B(A, \{\!|N_A \circ N_B|\!\}_{K_A})) \rightarrow \neg\mathsf{has}_{spy}(N_B)\,.$$

5. RELATED WORK

We now compare our work with related approaches to specifying and classifying security properties. Abadi and Tuttle [1] define a possible-worlds semantics for an extension of BAN that models interleaved protocol executions. However, details and examples are lacking so that a thorough comparison is difficult. Although their logic lacks an explicit notion of awareness, their *hide* operator conceals the contents of unreadable messages, and thus provides a basis for modeling "belief as a form of resource-bounded, defeasible knowledge" [1, p. 202]. It thereby captures some of the notions that our explicit belief modality formalizes.

In [2], we initially investigated how to use awareness to model resource-bounded reasoning in interleaved protocol executions. The multi-modal logic that we have given here differs considerably from [2]: while both are based on the explicit and implicit beliefs of the agents, here we modified and systematized the semantics for the modalities, the method how the awareness sets are computed, and how the logic is employed to specify properties and reason about attacks.

Interleaved trace-based semantics is a standard approach to modeling distributed computation. Paulson [18] has championed its use for inductive verification of security protocols, and the same semantic model can directly be used for model checking as well, e.g., as in [3]. Specifications in this setting (whether for verification or model checking) tend to be intensional as they are formalized in terms of sequences of protocol specific events. Our results in §4 illustrate how we can employ our modal specification to provide more abstract, high-level specifications of security properties with similar expressive power based on this semantic model.

Our definitions of intensional and extensional specifications come from Roscoe [19]. He also introduces the notion of *canonical intensional specification*, which "simply asserts that the protocol runs as expected" [19, p. 34], i.e. no agent can believe a protocol execution has completed unless the correct series of messages has occurred (consistent with all the various parameters) up to and including the last message the agent communicates. In our approach, this intensional character is directly formalized by the commit sets $\mathcal{C}$, and specified with the explicit belief modality. Note, however, that since we model action sets instead of action sequences, we cannot formalize the order in which the actions occur. However, it is straightforward to modify our framework to capture this idea.

A number of other authors, e.g. [5, 12, 17, 21], have looked at classifying and relating specifications. Notable in this regard is the work of Lowe [17], who uses CSP to formalize a hierarchy of authentication specifications, in which each level of the hierarchy expresses one possible meaning of "entity authentication". These specifications are all intensional; abstract notions such as secrecy are not accounted for. Using explicit belief it should be possible to formalize similar hierarchies in our setting. Moreover, using implicit belief it should be possible to extend these hierarchies, for example combining the intensional notion of "injective agreement" with the extensional requirement that some of the messages exchanged should remain secret.

6. CONCLUSIONS AND OUTLOOK

We have defined a multi-modal security logic with a trace-based semantics. Our logic combines the simple expressive semantics of trace-based approaches with the use of modalities to support high-level, trace-independent specifications of security properties based on different notions of belief. The logic also sheds light on, and allows us to give a logical characterization of, extensional and intensional specifications of security properties.

There is considerable work ahead and many interesting problems are still open. First, the account we have given is semantic. Via a semantic embedding, for example in higher-order logic, we could mechanize deductions in Isabelle (we have already carried out some initial work in this direction). More interesting though is to derive, from the semantics, higher-level proof rules for reasoning about the modalities.

Second, we have illustrated the logical equivalence between trace-based specifications (translated into our setting) and modal specifications. What is missing is a general statement about such equivalences. Such a statement is difficult as it requires the definition of a general class of trace-based specifications, and circumscribing such a class is problematic due to their intensional nature. One possible solution, which we would like to investigate, is to show equivalence for particular classes of specifications. For example, the semantics of the commit formulas captures an idea that is very close to the one of *matching histories* [8], except that, since we use sets of actions instead of sequences, we cannot talk about their ordering.

Finally, in our example in reasoning about attacks (i.e. the man-in-the-middle attack on the NSPK protocol) we knew of its existence in advance. One of the advantages of logics like BAN is that, in some cases, they allow for a kind of abductive reasoning as they provide a way of finding attacks by identifying missing assumptions required for proofs. When a deductive system for our logic is in place, we will also have the chance to explore these possibilities.

7. REFERENCES

[1] M. Abadi and M. R. Tuttle. A semantics for a logic of authentication. In *Proceedings of the 10th Annual ACM Symposium on Principles of Distributed Computing*, pages 201–216. ACM Press, 1991.

[2] R. Accorsi, D. Basin, and L. Viganò. Towards an awareness-based semantics for security protocol

analysis. In J. Goubault-Larrecq, editor, *Proceedings of the Post-CAV Workshop on Logical Aspects of Cryptographic Protocol Verification*, ENTCS 55(1). Elsevier, 2001.

[3] D. Basin. Lazy infinite-state analysis of security protocols. In R. Baumgart, editor, *Secure Networking: CQRE'99*, LNCS 1740, pages 30–42. Springer-Verlag, 1999.

[4] A. Bleeker and L. Meertens. A semantics for BAN logic. In *Proceeding of DIMACS Workshop on Design and Formal Verification of Crypto Protocols*. 1997.

[5] C. Boyd. Extensional goals in authentication protocols. In *Proceedings of DIMACS Workshop on Design and Formal Verification of Crypto Protocols*. 1997.

[6] M. Burrows, M. Abadi, and R. Needham. A logic of authentication. *ACM Transactions on Computer Systems*, 8(1):18–36, 1990.

[7] I. Cervesato and P. F. Syverson. The logic of authentication protocols. In R. Focardi and R. Gorrieri, editors, *Foundations of Security Analysis and Design*, LNCS 2171, pages 63–136. Springer-Verlag, 2001.

[8] W. Diffie, P. C. van Oorschot, and M. J. Wiener. Authentication and authenticated key exchanges. *Designs, Codes and Cryptography*, 2:107–125, 1992.

[9] D. Dolev and A. C. Yao. On the security of public key protocols. *IEEE Transactions on Information Theory*, 2(29):198–208, March 1983.

[10] R. Fagin and J. Y. Halpern. Belief, awareness and limited reasoning. *Artificial Intelligence*, 34(1):39–76, 1987.

[11] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about knowledge*. MIT Press, 1995.

[12] D. Gollmann. Authentication – myths and misconceptions. In *Progress in Computer Science and Applied Logic*. Birkhäuser Verlag, 2001.

[13] L. Gong, R. Needham, and R. Yahalom. Reasoning about belief in cryptographic protocols. In *Proceedings of the 1990 IEEE Symposium on Security and Privacy*, pages 234–248. IEEE Computer Society Press, 1990.

[14] G. E. Hughes and M. J. Cresswell. *A new introduction to modal logic*. Routledge, 1996.

[15] H. J. Levesque. A logic of implicit and explicit belief. In *Proceedings of AAAI'84*, pages 198–202. 1984.

[16] G. Lowe. Breaking and fixing the Needham-Schroeder public-key protocol using FDR. In *Proceedings of TACAS'96*, LNCS 1055, pages 147–166. Springer-Verlag, 1996.

[17] G. Lowe. A hierarchy of authentication specifications. In *Proceedings of the 10th IEEE Computer Security Foundations Workshop: CSFW'97*, pages 31–43. IEEE Computer Society Press, 1997.

[18] L. C. Paulson. The inductive approach to verifying cryptographic protocols. *Journal of Computer Security*, 6:85–128, 1998.

[19] A. W. Roscoe. Intensional specifications of security protocols. In *Proceedings of the 9th IEEE Computer Security Foundations Workshop: CSFW'96*, pages 28–38. IEEE Computer Society Press, 1996.

[20] P. F. Syverson. Knowledge, belief, and semantics in the analysis of cryptographic protocols. *Journal of Computer Security*, 1:317–334, 1992.

[21] P. F. Syverson and P. C. van Oorschot. On unifying some cryptographic protocol logics. In *Proceedings of the 1994 IEEE Computer Society Symposium on Research in Security and Privacy*. IEEE Computer Society Press, 1994.

[22] E. Thijsse. On total awareness logics. In M. de Rijke, editor, *Defaults and Diamonds*, pages 309–347. Kluwer Academic Publishers, 1993.

APPENDIX

PROOF OF LEMMA 8. We begin by observing that, for a formula $\varphi \in \mathcal{F}_A(w)$ such that $\varphi \in \{\mathsf{says}_A(B, M), \mathsf{sees}_A(M), \mathsf{has}_A(M)\}$, we trivially have that

$$\mathfrak{M}, w \models \varphi \text{ iff } \mathfrak{M}, w \models_A \varphi \tag{4}$$

since the definitions of $\mathfrak{M} \models \varphi$ and $\mathfrak{M} \models_A \varphi$ are in this case identical. The proof then proceeds as follows.

(Left-to-right) Assume that $\mathfrak{M} \models \mathcal{B}_A\, \varphi$. By definition, $\mathfrak{M}, w' \models \varphi$ for all w' such that $w \sim_A w'$, and thus $\mathfrak{M}, w \models \varphi$ as $w \sim_A w$ by definition. From (4), we then have that $\mathfrak{M}, w \models_A \varphi$, and thus, by definition, $\mathfrak{M}, w \models \mathcal{X}_A\, \varphi$.

(Right-to-left) Assume that $\mathfrak{M}, w \models \mathcal{X}_A\, \varphi$. By definition, $\mathfrak{M}, w \models_A \varphi$ and thus from (4) we have that $\mathfrak{M}, w \models \varphi$. Since φ characterizes a property of A's local state, if $\mathfrak{M}, w \models \varphi$ then $\mathfrak{M}, w' \models \varphi$ for all worlds w' in the equivalence class induced by A's indistinguishability relation $\sim_A$. Thus, by definition, $\mathfrak{M}, w \models \mathcal{B}_A\, \varphi$. □

PROOF OF THEOREM 9. We first show that $\mathfrak{M}^t_{\text{MITM}}, w_5 \models \mathcal{X}_b\, \mathsf{commR}_b(a, n_b)$ and $\mathfrak{M}^t_{\text{MITM}}, w_5 \not\models \mathcal{B}_b\, \mathsf{sec}_{\{a,b\}}(n_b)$. By definition of explicit belief, $\mathfrak{M}^t_{\text{MITM}}, w_5 \models \mathcal{X}_b\, \mathsf{commR}_b(a, n_b)$ iff $\mathfrak{M}^t_{\text{MITM}}, w_5 \models_b \mathsf{commR}_b(a, n_b)$ and $\mathsf{commR}_b(a, n_b) \in \mathcal{F}_b(w_5)$. From $Po_b(w_5)$ it follows that $\mathsf{commR}_b(a, n_b) \in \mathcal{F}_b(w_5)$ holds. By definition, $\mathfrak{M}^t_{\text{MITM}}, w_5 \models_b \mathsf{commR}_b(a, n_b)$ holds iff $\mathcal{C}_R^{\text{NSPK}}(a, b, n_b) \subseteq Ac_b(w_5)$, which holds because $\mathcal{C}_R^{\text{NSPK}}(a, b, n_b) = \{\mathtt{rec}_b(\{\!|a \circ n_a|\!\}_{k_b}), \mathtt{send}_b(a, \{\!|n_a \circ n_b|\!\}_{k_a}), \mathtt{rec}_b(\{\!|n_b|\!\}_{k_b})\}$. We can thus conclude that $\mathfrak{M}^t_{\text{MITM}}, w_5 \models \mathcal{X}_b\, \mathsf{commR}_b(a, n_b)$.

To show that b does not implicitly believe in $\mathsf{sec}_{\{a,b\}}(n_b)$ at w_5, observe that by definition $\mathfrak{M}^t_{\text{MITM}}, w_5 \not\models \mathcal{B}_b\, \mathsf{sec}_{\{a,b\}}(n_b)$ iff $\mathfrak{M}^t_{\text{MITM}}, w' \not\models \mathsf{sec}_{\{a,b\}}(n_b)$ for some w' such that $w_5 \sim_b w'$. Since w' can only be w_5 by the definition of $\sim_b$ in $\mathfrak{M}^t_{\text{MITM}}$, we check whether $\mathfrak{M}^t_{\text{MITM}}, w_5 \models \mathsf{sec}_{\{a,b\}}(n_b)$, which holds iff $\mathfrak{M}^t_{\text{MITM}}, w_5 \models \mathsf{has}_C(n_b)$ for all agents $C \in \{a, b\}$, and $\mathfrak{M}^t_{\text{MITM}}$, $w_5 \not\models \mathsf{has}_D(n_b)$ for all agents $D \notin \{a, b\}$. $Po_a(w_5)$ and $Po_b(w_5)$ tell us that both a and b possess n_b. Since we are only considering agents a, b and the *spy*, D can only be the *spy*. Since $\mathfrak{M}^t_{\text{MITM}}, w_5 \models \mathsf{has}_{spy}(n_b)$, we conclude that $\mathfrak{M}^t_{\text{MITM}}, w_5 \not\models \mathsf{sec}_{\{a,b\}}(n_b)$. □

PROOF OF THEOREM 10. (Left-to-right) We assume $\mathfrak{M}^t \models \mathcal{X}_B\, \mathsf{commR}_B(A, N_B) \rightarrow \mathcal{B}_B\, \mathsf{sec}_{\{A,B\}}(N_B)$ and $\mathfrak{M}^t, w \models \mathsf{sees}_B(\{\!|N_B|\!\}_{K_B}) \wedge \mathsf{says}_B(A, \{\!|N_A \circ N_B|\!\}_{K_A})$ for an arbitrary w, and show that $\mathfrak{M}^t, w \models \neg\mathsf{has}_{spy}(N_B)$.

$\mathfrak{M}^t, w \models \mathsf{sees}_B(\{\!|N_B|\!\}_{K_B})$ implies that there exists a message M such that $\mathtt{rec}_B(M) \in Ac_B(w)$ and $\{\!|N_B|\!\}_{K_B} \in \mathsf{submsg}_B(w, M)$. By the inductive definition of the protocol, M can only be $\{\!|N_B|\!\}_{K_B}$, which implies that $\mathtt{rec}_B(\{\!|N_B|\!\}_{K_B}) \in Ac_B(w)$.

$\mathfrak{M}^t, w \models \mathsf{says}_B(A, \{\!|N_A \circ N_B|\!\}_{K_A})$ implies that there exists an M such that $\mathtt{send}_B(A, M) \in Ac_B(w)$ and $\{\!|N_A \circ N_B|\!\}_{K_A} \in \mathsf{comp}_A(w, M)$. By the inductive definition of the protocol, M can only be $\{\!|N_A \circ N_B|\!\}_{K_A}$, which implies that $\mathtt{send}_B(A, \{\!|N_A \circ N_B|\!\}_{K_A}) \in Ac_B(w)$.

From $\mathtt{send}_B(A, \{\!|N_A \circ N_B|\!\}_{K_A}) \in Ac_B(w)$ and from $\mathtt{rec}_B(\{\!|N_B|\!\}_{K_B}) \in Ac_B(w)$ it follows, again by the inductive definition of the protocol, that $\mathtt{rec}_B(\{\!|A \circ N_A|\!\}_{K_B}) \in Ac_B(w)$. Thus, $\mathcal{C}_R^{\text{NSPK}}(A, B, N_B) \subseteq Ac_B(w)$. This implies that $\mathfrak{M}^t, w \models_B \mathsf{commR}_B(A, N_B)$. Since it is straightforward to show that $\mathsf{commR}_B(A, N_B) \in \mathcal{F}_B(w)$, we have that $\mathfrak{M}^t, w \models \mathcal{X}_B\, \mathsf{commR}_B(A, N_B)$.

The assumption and $\mathfrak{M}^t, w \models \mathcal{X}_B\, \mathsf{commR}_B(A, N_B)$ imply $\mathfrak{M}^t, w \models \mathcal{B}_B\, \mathsf{sec}_{\{A,B\}}(N_B)$, i.e. $\mathfrak{M}^t, w' \models \mathsf{sec}_{\{A,B\}}(N_B)$ for all w' such that $w \sim_B w'$. By the reflexivity of $\sim_B$ we have $\mathfrak{M}^t, w \models \mathsf{sec}_{\{A,B\}}(N_B)$, and by the definition of secrecy we have $\mathfrak{M}^t, w \not\models \mathsf{has}_C(N_B)$ for all $C \notin \{A, B\}$, so we can conclude that $\mathfrak{M}^t, w \models \neg\mathsf{has}_{spy}(N_B)$.

(Right-to-left) We assume $\mathfrak{M}^t \models (\mathsf{sees}_B(\{\!|N_B|\!\}_{K_B}) \wedge \mathsf{says}_B(A, \{\!|N_A \circ N_B|\!\}_{K_A})) \rightarrow \neg\mathsf{has}_{spy}(N_B)$ and $\mathfrak{M}^t, w \models \mathcal{X}_B\, \mathsf{commr}(B, A, N_B)$ for an arbitrary w, and show that $\mathfrak{M}^t, w \models \mathcal{B}_B\, \mathsf{sec}_{\{A,B\}}(N_B)$, i.e. $\mathfrak{M}^t, w' \models \mathsf{sec}_{\{A,B\}}(N_B)$ for all $w' \in \mathfrak{M}^t$ such that $w \sim_B w'$.

$\mathfrak{M}^t, w \models \mathcal{X}_B\, \mathsf{commR}_B(A, N_B)$ implies, by definition, that $\mathfrak{M}^t, w \models_B \mathsf{commR}_B(A, N_B)$. From the definition of commitment, it follows that $\mathcal{C}_R^{\text{NSPK}}(A, B, N_B) \subseteq Ac_B(w)$. Since $w \sim_B w'$, we have that $Ac_B(w) = Ac_B(w')$ and thus $\mathcal{C}_R^{\text{NSPK}}(A, B, N_B) \subseteq Ac_B(w')$. This implies both (i) $\mathfrak{M}^t, w' \models \mathsf{says}_B(A, \{\!|N_A \circ N_B|\!\}_{K_A})$, since $\mathtt{send}_B(A, \{\!|N_A \circ N_B|\!\}_{K_A}) \in Ac_B(w')$ and $\{\!|N_A \circ N_B|\!\}_{K_A} \in \mathsf{comp}_B(w', \{\!|N_A \circ N_B|\!\}_{K_A})$, and (ii) $\mathfrak{M}^t, w' \models \mathsf{sees}_B(\{\!|N_B|\!\}_{K_B})$, since $\mathtt{rec}_B(\{\!|N_B|\!\}_{K_B}) \in Ac_B(w')$ such that $\{\!|N_B|\!\}_{K_B} \in \mathsf{submsg}_B(w', \{\!|N_B|\!\}_{K_B})$. The assumption, together with (i) and (ii), implies $\mathfrak{M}^t, w' \models \neg\mathsf{has}_{spy}(N_B)$. Since B sent a message in the form of step two to A, we also have that $\{\!|N_A \circ N_B|\!\}_{K_A} \in Po_A(w')$ and $N_B \in \mathsf{analz}(Po_A(w'))$, which implies that $\mathfrak{M}^t, w' \models \mathsf{has}_A(N_B)$. Moreover, $\{\!|N_B|\!\}_{K_B} \in Po_B(w')$ and $N_B \in \mathsf{analz}(Po_B(w'))$, imply that $\mathfrak{M}^t, w' \models \mathsf{has}_B(N_B)$. It thus follows that $\mathfrak{M}^t, w' \models \mathsf{sec}_{\{A,B\}}(N_B)$ for an arbitrary w' such that $w \sim_B w'$, and hence $\mathfrak{M}^t, w \models \mathcal{B}_B\, \mathsf{sec}_{\{A,B\}}(N_B)$. □